Research on Face Expression Recognition based on Deep Learning and Machine Learning

Chenhao Si

University of Connecticut, USA

Keywords: Facial expression recognition; Deep learning; CNN; Machine learning; Computer vision; Image preprocessing; Feature extraction; Feature classification.

Abstract: The existing facial expression recognition technology is basically limited to the traditional machine learning algorithm. In the case of light intensity, occlusion and attitude transformation, the traditional machine learning algorithm has poor robustness and is difficult to apply to real life. With the development of hardware conditions such as computer GPU and the arrival of the era of big data, in-depth learning has attracted much attention in the field of computer vision. The traditional machine learning algorithms and their advantages and disadvantages are introduced in terms of processing, feature extraction and feature classification. The deep learning algorithms are introduced from DBN, CNN and other mainstream algorithms, development directions, common development frameworks. Finally, the development problems and trends of traditional machine learning and deep learning in facial expression recognition are summarized and prospected, as well as the future research directions.

1. Introduction

Face expression recognition, as an important branch of face recognition, has become a research hotspot in the field of human-computer interaction. It involves various disciplines, such as psychology, statistics, biology, computer science and so on. It is a relatively new and promising research direction. Its application is also very extensive. Many of them have been extended to the market and entered the lives of ordinary people. For example, fatigue driving, network teaching, human-computer interaction, medical field, entertainment industry. Psychologist Mehrabiadu [1] pointed out that facial expressions convey 55% of the useful information in the process of communication, while only 7% of the useful information conveyed by language. Computer facial expression recognition originated in the late 1970s, Suwa et al. will be sequential [2]. Face images are marked with 20 feature points and tracked for recognition. By comparing with the original model to realize facial expression recognition, Paul Ekman puts forward six basic expressions. They are happiness, sadness, surprise, anger, aversion, fear. Then he establishes the mapping relationship between different facial muscle movements and corresponding facial expressions through a lot of research. A facial action coding system (FACS) based on 44 motion units is proposed [3]. With the advancement of computer technology, the arrival of big data era and the development of computer hardware such as GPU, facial expression recognition has made great progress in both hardware and software, and more and more research institutions, expression databases and new algorithms have been developed. In ILSVRC, Hinton and his student Alex [4] won the championship of the International Computer Vision Competition. The recognition rate was raised from 74.3% to 84.7%. The breakthrough and substantial progress in the field of image recognition was realized. Since then, more scholars have joined the research. At present, the depth learning algorithm has improved the accuracy of the ImageNet dataset to 95.06% (the recognition rate of human being on the ImageNet data set is above 94.9% [5]). This recognition rate has been equal to the human recognition ability or even exceeds the human recognition rate. Different facial expression databases can affect the results of facial expression recognition. With the development of facial expression recognition, more and more facial expression databases are available. So far, the main databases are JAFFE [6], BHU [7], CK+ [8], MMI [9], BU-3DFE [10], GEMEP [11], FER2013[12].

2. Traditional Machine Learning Algorithms

Facial expression recognition based on traditional machine learning can be divided into three parts: image preprocessing, feature extraction, feature classification, image preprocessing, feature extraction and feature classification.

2.1 Image Preprocessing

Picture preprocessing can directly affect the extraction of feature points and the result of facial expression classification, thus affecting the recognition rate of facial expressions. For various reasons, pictures are usually polluted by some other signals, such as salt and pepper noise (impulse noise), Gauss noise, gamma noise, exponential distribution noise, uniform distribution noise, and so on. There is complex background, illumination intensity, occlusion, age and other interference factors. Moreover, the size of many databases is different, some are color images, some are gray images, and the quality of photographic equipment is uneven. These objective interference factors need to be preprocessed before recognition. Usually, the most important process of image preprocessing is shown in Fig. 1.

Fig. 1 Image preprocessing process

For an input image, first of all, denoising methods are average filter (AF), median filter (MF), adaptive median filter (AMF), Wiener filter (WF), etc. Face detection has developed into an independent research direction[13], while scale and gray normalization are from the following aspects: average filter (AF), median filter (MF), adaptive median filter (AMF), adaptive median filter (WF).The color information, size and other aspects of the image are processed to reduce the computational complexity while guaranteeing the key features of the face. Finally, histogram equalization enhances the image effect.

2.2 Feature Extraction

Feature extraction is the core part of the whole facial expression recognition system. How to extract the useful information of facial expression on the premise of guaranteeing the original information of the picture can improve the recognition rate of facial expression to a great extent. However, the dimension of most feature extraction is too high, which leads to dimension disaster. Therefore, it is usually necessary to reduce the dimension of features and decompose the features. At present, various feature extraction algorithms have been proposed and improved.

2.2.1 Gabor Feature Extraction

Gabor [14] wavelet kernel function has the same characteristics as the two-dimensional reflector of simple cells in human cerebral cortex, which can better describe texture features. It was first proposed by D. Gabor. Based on Fourier transform, because Gabor transform can't focus, researchers combine wavelet theory with Gabor feature to develop Gabor wavelet transform. Yu et al. [15] used a new algorithm of linear and non-linear synthesis on the basis of Gabor feature. Liu [16] In order to solve the problem that Gabor cannot acquire local expression features; a multi-directional feature fusion algorithm based on two Gabors is proposed. Wen [17] first selected a local small area artificially and extracted Gabor wavelet coefficients on this area, then calculated its mean value and used it as texture feature.

2.2.2 Local Binary Algorithms

Local binary pattern (LBP) [18] calculates the order of brightness between each pixel in the image and its local neighborhood points, and then encodes the binary order to form a local binary pattern. Finally, the multi-region histogram is used as the feature description of the image. The improved

LBP-based algorithm mainly includes the complete local binary algorithm (CLBP) proposed by Guo [19]. Based on the Fisher criterion, he improved the LBP operator (FCLLBP) [20]. Feng [21] achieved good recognition rate by establishing local LBP histogram and connecting it to recognize facial expressions. Zhou Yuxuan [22] improved LBP algorithm can select the unique features of a single expression to better recognize facial expression. The LBP algorithm is further improved to LDP (Local Orientation Mode) by Jabid and Chae [23]. This algorithm has good robustness to illumination and relatively low computational complexity. Local Phase Quantization [24] (LPQ) is mainly based on short-time Fourier Transform and has stable performance in feature extraction.

2.2.3 ASM/AAM

Active shape model (ASM) is proposed by Cootes [25]. This algorithm is based on statistical model and is generally used to extract feature points on facial contour. This model mainly uses global shape model to match the initial shape of face, and then establishes local texture model to obtain the contour features of target accurately [26]. Active appearance model (AAM) is based on ASM. The algorithm overcomes the disadvantage of using global features in ASM, and integrates local texture features into ASM. At the same time, it uses shape and texture features. Cristinacce [27] combines PRFR (feature response enhancement operator) with AAM to detect features of local edges such as facial organs. Hou Jie [28] improves ASM algorithm based on facial expression organs and locates 118 faces. Saatci [29] skillfully cascades AAM and Support Vector Machine (SVM) classifiers to improve recognition rate.

2.2.4 Haar-like Feature Extraction

The Haar-like feature is proposed by Viola. As shown in Fig. 2, the algorithm is a feature model composed of several rectangular regions with the same size and shape. These regions are defined as the sum of the pixel values of the black region and the sum of the pixel values of the white region [30]. Xu [31] combines the Haar-like and Adaboost algorithms. Firstly, Haar-like is used to get the most original features and calculate the eigenvalues of the whole picture. Then, Adaboost is used to train, extract different features and construct feature subspace.

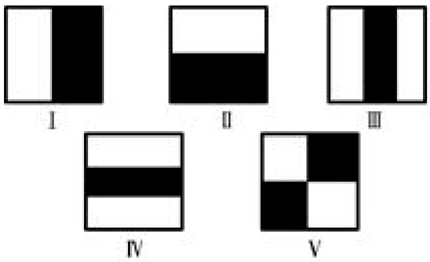

Fig. 2 Haar-like features

2.2.5 Optical Flow Method

Optical flow method [33] is the motion of image brightness/gray level in image sequence, that is, the expression of the velocity of points on the surface of space object on the imaging plane of vision sensor. Using Horn-Schunck (H-S) optical flow, the two-dimensional velocity field is combined with gray level, and the feature of moving object is extracted from continuous moving face image sequence. Zhao Xiaojian [34] proposed to solve the difficulty of feature extraction of moving object. An unsupervised behavior feature extraction method (DOF-SC) which combines dense optical flow trajectory with sparse coding framework is proposed. The algorithm sampled the original image centered on the DOF trajectory as the original feature of the trajectory. The algorithm does not need to label the feature points manually, so it reduces the complexity and uses sparse matrix to represent the feature. In theory, dimension reduction is not necessary, which reduces the load of the computer.

2.2.6 Feature Point Tracking Method

Yun [35] automatically extracts more than 20 points from the video stream as feature points of the face model, then uses particle filter to track these feature points, and constructs a variable 3D expression recognition model. Liu Yu et al. [36] proposed a feature point tracking method which combines KLT and SIFT. By improving SIFT, the feature points are evenly distributed and there is no clustering phenomenon. Then the KLT matching algorithm is designed hierarchically and iteratively, which can track and match quickly under the condition of obvious change of target attitude and size.

2.3 Feature Classification Algorithms

After feature extraction, in order to reduce the computational complexity of computer, feature dimension reduction is needed when the dimension is too high. PCA dimension reduction method and LDA dimension reduction method are commonly used. Another key to affect the recognition rate of facial expressions is how to select a suitable classifier to classify the features. The classifier can successfully predict the common feature scores of adult facial expressions and facial expressions recognition. Class algorithms include K-NN classification algorithm, SVM classification algorithm, Adaboost classification algorithm, Bayesian classification algorithm, hidden Markov model classification algorithm, DBN classification algorithm, etc.

2.3.1 K-NN Algorithm

K is an adjustable parameter in the nearest neighbor algorithm. Whether the computer can find the most suitable K value is the key to the classification performance. K-NN [32] classifier searches K nearest neighbor sample sets from the test data and takes the trained label as the sample label, which has the attributes on the sample. K-NN is an algorithm that directly classifies unknown data set labels and lets the computer "name" the label itself. The algorithm is simple and easy to implement, and the required sample set is not too large, but the classification speed of the algorithm is slow. When new samples are added, it is compared with the training set, and the weight of each attribute is the same, which results in a certain degree of reduction in classification accuracy.

2.3.2 SVM Algorithm

Support Vector Machine (SVM) is an algorithm that divides two classes of classes into hyperplane vector points on support plane. SVM [30] is based on VC dimension theory and structural risk minimization principle. By providing limited sample data information, we can find the best compromise scheme on complex models and obtain the best generalization ability. Through the kernel function, we can map the linearly inseparable data to a higher dimension and transform it into linearly separable ones. By introducing kernels, computers can effectively process high-dimensional data and avoid dimension disaster to a certain extent.

2.3.3 Adaboost Algorithm

Adaboost [41] algorithm is a classical Two-Classifier algorithm formed by Yoav Freund and Rober Schapire. It finds the best weak classifier for the weight distribution of the training sample set by iteration. According to the weak classifier, the weight parameters of the wrong samples are adjusted. After meeting the iteration times of threshold, several weak classifiers are combined into a strong classifier. Also, it can be used together with other algorithm and will make the work more effective. Adaboost algorithm can work on the outliers and the irrelevant variables. Thus, it will make fewer mistakes in such scenario.

2.3.4 Bayesian Classification

Bayesian network is a probabilistic network based on the statistics called Baye’s Theorem [42]. So, it works based on statistics. The key idea of the algorithm is to graphicalize the network by probabilistic reasoning. It can give us more precise prediction of whether the candidates belong to this category or not by using probabilities. And because of the use of probability, we can get the graph model with the Bayesian network. Bayesian formula and Bayesian algorithm are often used in target recognition, tracking, medical diagnosis, biology, coding, etc.

2.4 Intrinsic Linkages and Development Trends of Algorithms

Traditional machine learning algorithms are different and related. Gabor wavelet has good robustness to multi-scale and multi-direction texture feature transformation. Compared with other algorithms, it can detect the change better and is insensitive to the intensity of illumination. But Gabor wavelet usually acts on global features, so it consumes more memory. Compared with LBP operator, it has less storage space and higher computational efficiency, but LBP operator has higher computational efficiency. The operator is inefficient to extract features from noisy images, and only considers the pixel features of the image center and neighborhood, ignoring the difference magnitude will lead to the loss of some useful feature information. For expression feature extraction, Harr has proved to be more advantageous because it describes the local gray level change of the face. When the global region light is stable, Harr can extract more facial motion unit changes. For AAM and ASM, AAM is based on ASM. ASM has higher execution efficiency, AAM is relatively complex, can better fit texture features, recognition rate is higher than ASM. In feature classification, SVM is used more widely. SVM usually trains the model in training set first. Then the trained model is used to classify the test set. KNN has no training process and only measures the Euclidean distance between training data. Adaboost solves the problem of low training accuracy caused by insufficient training samples by resampling training samples. Usually, Adaboost is used to enhance the classification ability of SVM. The working mechanism of several classification algorithms is similar [43]. As shown in Fig. 3, no matter which classification algorithm is based on machine learning a classification boundary, only Adaboost can enhance the classification ability of SVM. Some are relatively complex ("+, -" representing two categories).

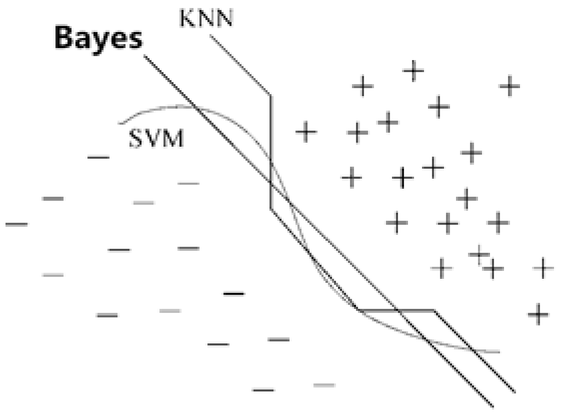

Fig. 3 Different classification criteria can produce the same prediction results

Table 1 lists the features, classification algorithms and their recognition rates and databases used by some traditional machine learning algorithms. Traditional machine learning algorithms have been developed continuously and are relatively mature. The following research directions are mainly towards micro-expressions [33] and negative expressions. How to quickly capture the changes of facial microexpressions and solve the problem of low recognition rate of negative expressions need to be explored by researchers.

3. Deep Learning Algorithms

In 2016, AlphaGo played a game with Li Shishi. At last, AlphaGo defeated Li Shishi by 4:1. The man-machine war showed that machine had defeated human brain to some extent. At present, the traditional machine learning method has been unable to meet the needs of processing time, performance and intelligence in the era of big data, while in-depth learning has shown excellent information processing ability, especially in sub-division. Class recognition and target detection have become a research hotspot both at home and abroad since 2012. They are one of the ten innovative technologies that changed the world in Science American in 2015[44]. In facial expression recognition, in-depth learning avoids tedious image preprocessing and feature extraction, and performs better than traditional methods. It has better robustness to facial expression recognition such as illumination, posture and occlusion.

Table 1. Part 1 Machine Learning Algorithms Recognition Results

Image	Reference	Feature method	Classification method	Recognition rate /%	Data base
Dynamic image	[38]	LBP / VAR	DBN model	91.40	CK+
	[37]	ASM	Classification Rules Based on AU Block	83.60	Autonomous collection
	[33]	H-S Optical Flow Method	Global Threshold Algorithms	80.70	SMIC
	[35]	AAM Combined with ASM	D-ISOMAP	88.20	RML
	[34]	DOF-SC	SVM	90.20	YouTube Library
	[39]	Gabor wavelet	FBMM	94.90	CK
	[53]	Zemike moments	NB	73.20	CK
Still image	[40]	Gabor+LBP	RBF	93.42	JAFFE
	[32]	HOG	AdaBoost	88.70	CK+
	[30]	Haar-Like	SVM	98.10	CK
	[23]	LDP	RBF	90.40	JAFFE
	[23]	LDP	RBF	93.40	CK
	[32]	HOG	SVM	89.50	JAFFE
	[22]	DiscLBP	KNN	97.30	CK
	[41]	Gabor	GPC + AdaBoost	90.90	JAFFE
	[41]	Gabor	GPC+AdaBoost	91.50	CK
	[31]	CA	HMM	95.80	JAFFE
	[42]	Gabor	KNN+Bayesian Classification	92.00	CK+

3.1 CNN

CNN (convolutional neural network) is an end to end model and an improvement of artificial neural network (ANN). It was inspired by the study of the cat's primary visual cortex; simple cells respond by receiving stimuli from specific edges of local receptive fields as inputs to complex cells. Its network structure is shown in Fig. 4. The most important characteristics of CNN are local connectivity and weight sharing, which reduces network parameters, speeds up training and has a certain regularization effect. CNN is a complex neural network. Loss function [50] is an evaluation criterion for evaluating the coincidence between predicted and real values. The commonly used loss functions are mutual entropy loss of Softmax classifier, hinge loss, L2 norm loss and L1 norm loss. Norm loss CNN can solve the problem of nonlinearity by activation function [51], which preserves and maps the characteristics of activated neurons. The commonly used activation functions are tanh, sigmoid, soft sign, ReLU, LReLU, etc. In order to prevent the over-fitting phenomenon caused by too deep network layers and increase its generalization ability, CNN generally needs regularization operation Regularization method commonly used. On the other hand, CNN uses pooling [52] (pooling layer) operation to minimize network complexity and keep CNN locally invariant, thus closer to the mechanism of animal visual cortex.

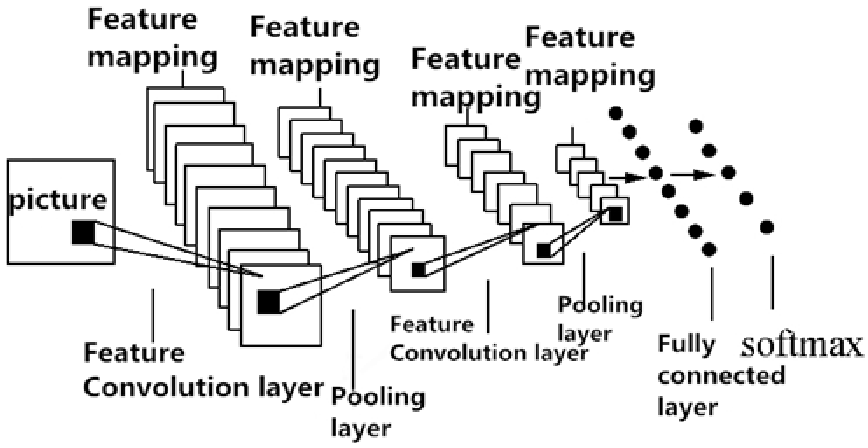

Fig. 4 CNN Network Architecture

3.2 DBN

Deep belief network (DBN) is based on restricted Boltzmann machine (RBM). It consists of a backward feedforward propagation network and a deep RBM. The RBM model takes the front output as the back input and carries out multiple RBM layers. Each layer of RBM contains a hidden layer and a visible layer. Unlike BM (Boltzmann machine), RBM does not connect each layer. The network structure of the hidden layer and the visible layer is like the cognitive process of the human brain structure. Its feature of extracting input signals is unsupervised and abstract layer by layer [38]. DBN is a self-learning process from low to high level. It has the characteristics of initiative and accuracy. The disadvantage is that the convergence speed is slow and it is easy to converge to local optimum.

3.3 Major Algorithms Optimization and Development Direction

After several years of development, deep learning has achieved some results in the field of expression recognition. Yu [54] constructs a nine-layer CNNs structure. In the last layer, the expression is classified into seven categories by using soft Max classifier. The recognition rate of this model on the SFEW 2.0 data set reaches 61.29%. Lopes [55] joins the pretreatment process before the CNN network to explore the effect of pretreatment on the accuracy. The final recognition rate is in CK+. Wang [56] trained CNN model with soft max, adjusted parameters with triple loss function, and enhanced recognition rate by 2%. The model performed well on indistinguishable expressions (such as anger and aversion). Zhao [57] combines MLP and DBN, linking the advantages of unsupervised feature learning of DBN with the classification advantages of MLP to improve performance. He [38] combines deep learning with traditional machine learning. First, LBP/VAR is used to extract initial features and classify them as input of DBN. Li [58] in order to solve the problem that DBN ignores local features of images, CS-LBP and DBN are integrated. Deep learning has developed so far, various models and variants have been proposed, and the basic models cannot meet the performance requirements. Based on a large number of literature data, the future development direction of facial expression recognition is mainly reflected in improving network depth, hybrid model, combination with traditional machine learning, migration learning, unsupervised learning and so on. At the same time, the development of mobile terminals has begun to attract attention.

3.4 Common Application Framework

A good development framework can speed up the development of in-depth learning. At present, several mainstream in-depth learning frameworks are: Theano [45] has a grandfather status in in in-depth learning. Developed in 2007, based on Python, it is very suitable for data exploration, but because it is too low-level to support distributed, more and more researchers are developing Tensorflow; Tensorflow [46] is developed by Google in 2015.Open source has become the most popular open source project for machine learning on GitHub in 1997. Its flexible, universal, visualized and mobile advantages make it suitable for both research and product development. Keras [47] is a new framework in the open source framework for deep learning. Its background is based on Theano / Tensorflow, written in Python, and it has the advantage of easy introduction; Caffe [48], which was developed in 2013, plays a very important role in the deep learning framework. Especially in the field of image classification, it represents the first-class standard in the research field. But it cannot be well applied to other deep learning applications, such as voice and word processing. Torch [49] is written in Lua language. Because most developers are unfamiliar with Lua, it is difficult to promote Lua, but its flexibility is very high and it is easy to compile. My own level, and there are many well-trained models.

4. Conclusion

Deep learning has become a research hotspot in many scientific fields such as computer vision. In the field of facial expression recognition, deep learning solves the sensitive problems of traditional machine learning such as facial posture, illumination, occlusion, and improves the robustness of facial expression recognition. Traditional research on facial expression recognition has turned to micro-expression, and some from 2D to 3D facial modeling to construct a three-dimensional facial expression recognition system. But this does not mean that traditional machine learning methods have been eliminated, combined with in-depth learning, traditional machine learning algorithms may be revitalized. At present, there are the following problems to be solved in traditional machine learning and in-depth learning

1) Deep learning is data-driven. Massive data sets with annotations are the cornerstone of deep learning algorithms. Because of millions of data with labels such as ImageNet, CNN can perform well in competitions such as ILSVRC. But in fact, it is particularly difficult to construct a massive and labeled database, and how to automatically label the massive unlabeled data is also a direction that needs attention.

2) Traditional machine learning algorithm has absolute advantage in small data sample analysis. Deep learning algorithm is prone to fit phenomenon in small data sample. How to compromise these two phenomena and create an algorithm that can perform well in small data sample and large data is a key problem to be solved urgently in face expression recognition research.

3) Although deep learning has excellent performance in the field of facial expression classification, its theoretical knowledge is not yet complete. Most scholars use the method of parameter to improve the recognition rate. They regard deep learning as a black box. How to use theoretical knowledge to guide practice and promote the understanding of theoretical knowledge through practice is what researchers need to solve.

4) Although the recognition efficiency of in-depth learning is very fast after training the model, it is a long process in the training stage. Even with hardware acceleration such as GPU, training the model under deeper network and more massive data often takes several days or longer for researchers.

5) The expression distinction is not detailed enough. At present, most expression databases are mainly composed of seven basic expressions: happiness, sadness, surprise, anger, disgust, fear and neutrality. Few databases contain more subtle expressions, which hinders human-computer interaction from moving towards a more intelligent way.

In conclusion, how to deal with and understand the relationship between in-depth learning and traditional machine learning, learn from each other, learn from others' strong points and close the gap, is the focus and direction of future research on facial expression recognition.

References

[1] MEHRABIAN A, RUSSELL J A. An approach to environmental psychology [M]. Cambridge, MA: MIT, 1974.

[2] SUWA M, SUGIE N, FUJIMORA K. A preliminary note on pattern recognition of human emotional expression [C]/ Proceedings of the 4th International Joint Conference on Pattern Recognition. Kyoto, Japan, 1978: 408-410.

[3] EKMAN P, FRIESEN W V, HAGERJ C. Facial action coding system (FACS) [M]. Manual, Salt Lake City (USA): A Human Face, 2002.

[4] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [C]/ Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, Nevada, 2012: 1097-1105.

[5] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Delving deep into rectifiers: surpassing human-level performance on Image Netclassification [C]/ Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile, 2015: 1026-1034.

[6] LYONSM J, KAMACHIM, GYOBA J. The Japanese Female Facial Expression (JAFFE) database [EB/OL]. http://www.kasrl.org/jaffe.html.

[7] Design and implementation of Xue Yuli, Maoxia, Zhangfan. BHU facial expression database [J]. Journal of Beijing University of Aeronautics and Astronautics, 2007, 33 (2): 224-228.

[8] LUCEY P, COHN J F, KANADE T, et al. The Extended Cohn-Kanade Dataset (CK+): a complete dataset for action unit and emotion-specified expression [C]/ Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Recognition-Workshops. San Francisco, CA, USA, 2010: 94-101.

[9] VALSTAR M F, PANTICM. Induced disgust, happiness and surprise: an addition to the MMI facial expression database [C]/ Proceedings of Int & apos; l Conference on Language Resources and Evaluation, Workshop on EMOTION. Malta, 2010: 65-70.

[10] YIN Lijun, WEI Xiaozhou, SUN Yi, et al. A 3D facial expression database for facial behavior research [C]/ Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition. Southampton, UK, 2006: 211-216.

[11] VALSTAR M F, MEHU M, JIANG Bihan, et al. Meta-analysis of the first facial expression recognition challenge [J]. IEEE transactions on systems, man, part B (cybernetics), 2012, 42 (4): 966-979.

[12] GOODFELLOW I J, ERHAN D, CARRIER P L, et al. Challenges in representation learning: a report on three machine learning contests [M]/ LEE M, HIROSEA, HOU Zengguan, et al. Neural Information Processing. Berlin Heidelberg: Springer, 2013.

[13] YANGM H, KRIEGMAN D J, AHUJA N. Detecting faces inimages: a survey [J]. IEEE transactions on pattern analysis and machine intelligence, 2002, 24 (1): 34-58.

[14] LEE T S. Image representation using 2D Gabor wavelets [J]. IEEE transactions on pattern analysis and machine intelligence, 1996, 18 (10): 959-971.

[15] YU Jiangang, BHANU B. Evolutionary feature synthesis for facial expression recognition [J]. Pattern recognition letters, 2006, 27 (11): 1289-1298.

[16] LIU Shuaishi, TIAN Yantao, WAN Chuan. Facial expression recognition method based on Gabor multi-orientation features fusion and block histogram [J]. Actaautomaticasinica, 2011, 37 (12): 1455-1463.

[17] WEN Zhen, HUANG T S. Capturing subtle facial motions in 3D face tracking [C]/ Proceedings of the 9th IEEE International Conference on Computer Vision. Nice, France, 2003: 1343-1350.

[18] AHONEN T, HADID A, PIETIK FINEN M. Face recognition with local binary patterns [M]/ PAJDLA T, MATASJ. Computer Vision-ECCV 2004. Berlin Heidelberg: Springer, 2004: 469-481.

[19] GUO Zhenhua, ZHANG Lei, ZHANG D. A completed model of local binary pattern operator for texture classification [J]. IEEE transactions on image processing, 2010, 19 (6): 1657-1663.

[20] GUO Yimo, ZHAO Guoying, PIETIK FINEN M. Discriminative features for texture description [J]. Pattern recognition, 2012, 45 (10): 3834-3843.

[21] FENG X, PIETIK FINEN M, HADID A. Facial expression recognition based on local binary patterns [J]. Pattern recognition and image analysis, 2007, 17 (4): 592-598.

[22] Zhou Yuxuan, Wu Qin, Liang Jiuzhen, et al. Discriminant Complete Local Binary Pattern Face Expression Recognition [J]. Computer Engineering and Application, 2017, 53 (4): 163-169, 194.

[23] JABID T, KABIR M H, CHAE O. Robust facial expression recognition based on local directional pattern [J]. ETRI journal, 2010, 32 (5): 784-794.

[24] LONARE A, JAIN S V. A survey on facial expression analysis for emotion recognition [J]. International Journal of advanced research in computer and communication engineering, 2013, 2 (12): 4647-4650.

[25] COOTES T F, TAYLOR C J, COOPER D H, et al. Active shape models - their training and application [J]. Computer vision and image understanding, 1995, 61 (1): 38-59.

[26] Huang Jian, Li Wen, Gao Yujuan. Advances in facial expression recognition [J]. Computer Science, 2016,43 (S2): 123-126.

[27] CRISTINACCE D, COOTES T, SCOTT I. A multi-stage approach to facial feature construction [C]/ Proceedings of the British Machine Vision Conference. Kingston, 2004: 231-240.

[28] Hou Jie. Research on Face Expression Computing Technology [D]. Suzhou: Suzhou University, 2014.

[29] SAATCI Y, TOWN C. Cascaded classification of gender and facial expression using active appearance models [C]/ Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition. Southampton, UK, 2006: 393-398.

[30] XU Chao, DONG CAICHAO, FENG ZHIYONG, et al. Facial expression pervasive analysis based on Haar-like features and SVM [M]/KHACHIDZEV, WANG T, SIDDIQUI, et al. Contemporary Research on E-business technology and Strategy. BerlinHeidelberg: Springer, 2012: 521-529.

[31] Zhou Shuren. Analysis and Research of Facial Expression Recognition Algorithms [D]. Changsha: Central South University, 2009.

[32] LIEW C F, YAIRI T. Facial expression recognition and analysis: a comparison study of feature descriptors [J]. IPSJ transactions on computer vision and applications, 2015, 7: 104-120.

[33] Jiang Bo, Jielun, Liu Xin, et al. Microexpression capture based on optical flow modulus estimation [J]. Journal of Zhejiang University: Engineering Edition, 2017, 51 (3): 577-583, 589.

[34] Zhao Xiaojian, Zeng Xiaoqin. Behavior recognition method based on dense optical flow trajectory and sparse coding algorithm [J]. Computer application, 2016,36(1): 181-187.

[35] TIE Yun, GUAN Ling. A deformable 3-D facial expression model for dynamic human emotional state recognition [J]. IEEE transactions on circuits and systems for video technology, 2013, 23 (1): 142-157.

[36] Liu Yu, Wang Jingdong, Li Peng. A feature point tracking method based on SIFT and KLT [J]. Journal of Astronautics, 2011, 32 (7): 1618-1625.

[37] SALAKANIDOU F, MALASSIOTIS S S. Real-time 2D + 3D facial action and expression recognition [J]. Pattern recognition, 2010, 43 (5): 1763-1775.

[38] HE Jun, CAI Jianfeng, FANG Lingzhi, et al. Facial expression recognition based on LBP/VAR and DBN model [J]. Application research of computers, 2016,33 (8): 2509-2513.

[39] ZHAN Yongzhao, CHENG Keyang, CHEN Yabi, et al. A new classifier for facial expression recognition: fuzzy buried Markov model [J]. Journal of computer science and technology, 2010, 25 (3): 641-650.

[40] Li Dong. Facial expression recognition method based on local texture feature fusion [D]. Changchun: Jilin University, 2014.

[41] Li Wen, He Fangfang, Qian Yuantao, et al. Facial expression recognition based on Adapoost-Gauss process classification [J]. Journal of Zhejiang University: Engineering Edition, 2012, 46 (1): 79-83.

[42] Jielun, Lu Yanan, Jiang Bo, et al. Automatic facial expression recognition based on facial motion unit and expression relationship model [J]. Journal of Beijing University of Technology, 2016, 36 (2): 163-169.

[43] DOMINGOS P. A few useful things to know about machine learning [J]. Communications of the ACM, 2012, 55 (10): 78-87.

[44] Liu Dong, Li Su, Cao Zhidong. In-depth learning and its application in image object classification and detection [J]. Computer Science, 2016, 43 (12): 13-23.

[45] BASTIEN F, LAMBLIN P, PASCANU R, et al. Theano: new features and speed improvements [J]. arXiv: 1211. 5590, 2012.

[46] ABADI M, BARHAM P, CHEN Jianmin, et al. TensorFlow: a system for large-scale machine learning [C]/ Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation. Savannah, GA, USA, 2016.

[47] SUI Xudong, ZHANG Jinfang, HU Xiaohui, et al. Monitoring target through satellite images by using deep convolutional networks [M]/ LEE R. Software Engineering, Management and Applications. Cham: Springer, 2016.

[48] JIA Yangqing, SHELHAMER E, DONAHUE J, et al. Caffe: convolutional architecture for fast feature embedding [J]. Arxiv: 1408.5093, 2014.

[49] BOJARSKI M, DEL TESTA D, DWORAKOWSKI D, et. al. End to end learning for self-driving cars [J]. arXiv: 1604.07316, 2016.

[50] CS231 nconvolutional neural network work for visual recognition [EB/OL]. http:// cs231n. github.io/neural-networks-2/.

[51] ZHAO Shichao, LIU Yanbin, HAN Yahong, et al. Pooling the convolutional layers in deep ConvNets for video action recognition [J]. arXiv: 1511.02126, 2015.

[52] Xue Kunnan. Research on visual recognition based on convolution neural network [D]. Guangzhou: South China Agricultural University, 2016.

[53] LAJEVARDI S M, HUSSAIN Z M. Higher order orthogonal moments for invariant facial expression recognition [J]. Digital signal processing, 2010, 20 (6): 1771-1779.

[54] YU Zhiding, ZHANG Cha. Image based static facial expression recognition with multiple deep network learning [C]/Proceedings of 2015 ACM on International Conference on Multimodal Interaction. Seattle, Washington, USA, 2015: 435-442.

[55] LOPES A T, DE AGUIAR E, OLIVEIRA-SANTOS T. A facial expression recognition system using convolutional networks [C]/ Proceedings of the 28th SIBGRAPI Conference on Graphics, Patterns and Images. Salvador, Brazil, 2015: 273-280.

[56] WANG Jieru, YUAN Chun. Facial expression recognition with multi-scale convolution neural network [M]/ CHEN Enqing, GONG Yihong, TIE Yun. Advances in Multimedia Information Processing-PCM 2016. Cham: Springer, 2016.

[57] ZHAO Xiaoming, SHI Xugan, ZHANG Shiqing. Facial expression recognition via deep learning [J]. IETE technical review, 2015, 32 (5): 347-355.

[58] LI Chen, WEI Wei, WANG Jingzhong, et al. Face recognition based on deep belief network combined with central symmetric local binary pattern [M]/ PARK J, JIN Hai, JEONG Y S, et al. Advanced Multimedia and Ubiquitous Engineering. Singapore: Springer, 2016.

[59] CAI jia, HUANG panfeng. Research of a Real-time FeaturePoint Tracking Method Based on the Combination of Improved SURF and P-KLT Algorithm [J]. Acta Aeronautica et Astronautica Sinica, 2013, 05.